

AD A095024

LEVEL

12

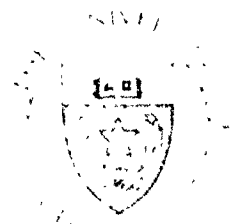
CENTER FOR
CYBERNETIC
STUDIES

The University of Texas
Austin, Texas 78712

DTIC
S FEB 17 1981
C

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited



81 2 17 056

12

Research Report CCS 387

MINIMIZATION BY INTERPOLATION:
A SECOND ORDER GRADIENT ALGORITHM

by

J. Barzilai

November 1980

NOV 1980
C

This research was partly supported by ONR Contract N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas at Austin
Austin, TX 78712
(512) 471-1821

DISTRIBUTION STATEMENT 1
Approved for public release
Distribution Unlimited

Newton's method for finding a stationary point of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ consists of the iteration

$$x_{i+1} = x_i - [\nabla^2 f(x_i)]^{-1} \cdot \nabla f(x_i) .$$

Its main attraction is its second order of convergence. However, it necessitates computation and inversion of the second order derivatives matrix.

Common minimization algorithms approximate the Hessian or its inverse by first order (i.e. gradient) information. First order information algorithms in common use, have at best superlinear rate of convergence [cf. 2].

We present a new class of algorithms which use first order information only, while maintaining quadratic convergence.

At step i of the algorithm, we interpolate f by a suitable interpolating function T , requiring

$$(1) \quad \begin{cases} T(x_{i-j}) = f(x_{i-j}) \\ \nabla T(x_{i-j}) = \nabla f(x_{i-j}) , \end{cases} \quad j = 0, 1$$

and determine x_{i+1} as a solution of the equation

$$(2) \quad \nabla T(x_{i+1}) = 0 .$$

We assume that the interpolating function depends on some parameters. We further assume that the equations (1) for the parameters of T , and equation (2) for x_{i+1} have solutions for all i , and that the parameters of T depend on the data continuously through (1). Finally, we assume that f and T have continuous derivatives of order 5 near the solution.

Approved for	
THIS CASE	
THIS TAB	
Unannounced	
Justification	
By	
Distribution	
Approved	

A

We derive the rate of convergence of the algorithm defined by (1) and (2) by establishing a difference relation for the errors $e_i = \|x_i - x^*\|$. Here $\|\cdot\|$ is an arbitrary fixed norm. This difference relation is analogous to the one obtained in [1] for the one-dimensional case.

To this end we define a function $\psi: R \rightarrow R^n$ in terms of which and the functions f, T we can express the errors of a related one-dimensional interpolation problem.

We assume that a point $x^* \in R^n$ which is a solution of

$$(3) \quad \nabla f(x^*) = 0$$

exists. Let $\psi: R \rightarrow R^n$ be a curve in R^n through the points x_{i-j} $j=1,0,-1$ and x^* , i.e.,

$$(4) \quad \begin{cases} \psi(t_{i-j}) = x_{i-j} & , \quad j=1,0,-1, \\ \psi(t^*) = x^* & , \end{cases}$$

where the parameter t is chosen so that

$$(5) \quad t_{i-j} = \|x_{i-j} - x^*\|, \quad t^* = \|x^* - x^*\| = 0.$$

We will later discuss the existence of this construction. Note, however, that the construction of ψ is a part of the analysis of the properties of the algorithm, not a part of the algorithm itself.

Now define $\theta(t) = T(\psi(t))$, $\phi(t) = f(\psi(t))$. Equations (1) and (4) imply

$$\theta^{(k)}(t_{i-j}) = \phi^{(k)}(t_{i-j}) \quad j, k = 0, 1,$$

which in turn implies [see 3]

$$(6) \quad \phi(t) - \theta(t) = \frac{\phi^{(4)}(\eta) - \theta^{(4)}(\eta)}{4!} \prod_{j=0}^3 (t - t_{i-j})^2.$$

where η is some intermediate point. Equation (16) is the basic difference relation we need (cf. [1]). Differentiating it and setting $t=0$, we obtain

$$(7) \quad t_{i+1} = B_i t_i t_{i-1}^2.$$

If the sequence B_i converges to a non-zero limit, the relation (7) implies that the sequence t_i converges to zero if t_0, t_1 are small enough, with rate of convergence which is given by the unique positive root of the indicial polynomial of (7): $t^2 - t - 2 = 0$, i.e. quadratically [cf. 4].

In order for the sequence B_i to converge, it is sufficient that $\theta^{(5)}$ and $\phi^{(5)}$ exist and are continuous, and $\phi''(0) \neq 0$. This would be the case if f has continuous derivatives of order 5 near the solution, the parameters of T depend continuously on the data, and T has continuous derivatives of order 5 for the appropriate values of the parameters. Finally, it is evident that the curve ψ can be chosen so that $\phi''(0) = \dot{\psi}' \nabla^2 f \dot{\psi}$ is nonzero, e.g. by choosing

$$\psi_k = a_k t^4 + b_k t^3 + c_k t^2 + t + d_k, \quad k = 1, \dots, n.$$

Note that no line search is needed in this class of algorithms, and that they may be designed to locate saddle points rather than minimum points.

A useful choice for the interpolating function T seems to be a separable sum of rational functions of the type discussed in [1].

The results in [1] for the one-dimensional case, can clearly be extended by the same device to the n -dimensional case. In particular, algorithms based on

function values only, have rates of convergence between 1.3 and 1.6; the rate of convergence is independent of the interpolating function, and inverse interpolation can be utilized for minimization. Similar results hold for the root-finding problem discussed by Traub [5]. Details of this work will appear elsewhere.

REFERENCES

- [1] J. Barzilai, "Convergent Methods for Optimization Problems," Ph.D. Dissertation, Technion, Israel Institute of Technology, Haifa, Israel, 1980.
- [2] J. Dennis and J. Moré, "Quasi-Newton Methods, Motivation and Theory," SIAM Rev., 19(1977), 46-87.
- [3] A.M. Ostrowski, Solution of Equations and Systems of Equations, Second Edition, Academic Press, 1966.
- [4] A. Tamir, "Rates of Convergence of a One-Dimensional Search Based on Interpolating Polynomials," JOTA, 27(1979), 187-203.
- [5] J.F. Traub, Iterative Methods for Solution of Equations, Prentice-Hall, 1964.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(9) Research rept.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
14 REPORT NUMBER CCS-387	2 GOVT ACCESSION NO. AD-A095 024	3 RECIPIENT'S CATALOG NUMBER	
4 TITLE (and Subtitle) Minimization by Interpolation: A Second Order Gradient Algorithm.		5 TYPE OF REPORT & PERIOD COVERED	
7 AUTHOR(s) Jonathan Barzilai		6 PERFORMING ORG REPORT NUMBER CCS 387	
		8 CONTRACT OR GRANT NUMBER(s) N00014-75-C-0569	
9 PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies The University of Texas at Austin Austin, TX 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11 CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 434) Washington, DC		12 REPORT DATE November 1980	
14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12/8		13 NUMBER OF PAGES 6	
		15. SECURITY CLASS. (of this report) Unclassified	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16 DISTRIBUTION STATEMENT (of this Report) This document has been approved for public release and sale; its distribution is unlimited.			
17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18 SUPPLEMENTARY NOTES			
19 KEY WORDS (Continue on reverse side if necessary and identify by block number) Unconstrained minimization, Interpolation, Convergence rates.			
20 ABSTRACT (Continue on reverse side if necessary and identify by block number) Newton's method for finding a stationary point of $f:R^h \rightarrow R$ has <u>second order</u> of convergence. However, it necessitates computation and inversion of the <u>second order</u> derivatives matrix. We present new classes of algorithms. One of these has <u>second order</u> of convergence while using <u>first order</u> information only.			

END

DATE
FILMED

7-80

DTIC